Facility Registry Service (FRS)

Automated and Manual Facility Merging

Date: September 2013

Overview

|	
Procedure Title	Automated and manual merging in FRS
Procedure Owner	U.S. EPA FRS
Document Number	
Version Number	1.1
Version Date	September 2013
Approved By	
Approved Date	
Comments	

TABLE OF CONTENTS

Purpose & Overview

The purpose of this documentation is to showcase the methodologies used for merging duplicate facilities within FRS. Facilities are merged automatically as well as manually. As FRS receives data flows from its partners, there are times when the incoming facility is already in the FRS database. In this case if the facility is an exact match or meets a certain criteria, it is automatically linked. If the facility is deemed as a potential duplicate, a flag is raised, it is then referred to the data stewards for further analysis and resolution (merge manually or clear flag). This merging capability and function reduces redundancy within the system, making the service more concise, and reduces burden of data management in the long run.

Automated Merging

The automated merging is done by PL/SQL logic in the Oracle database.

When a new record is inserted into FRS it follows the logic shown in the flowchart below:
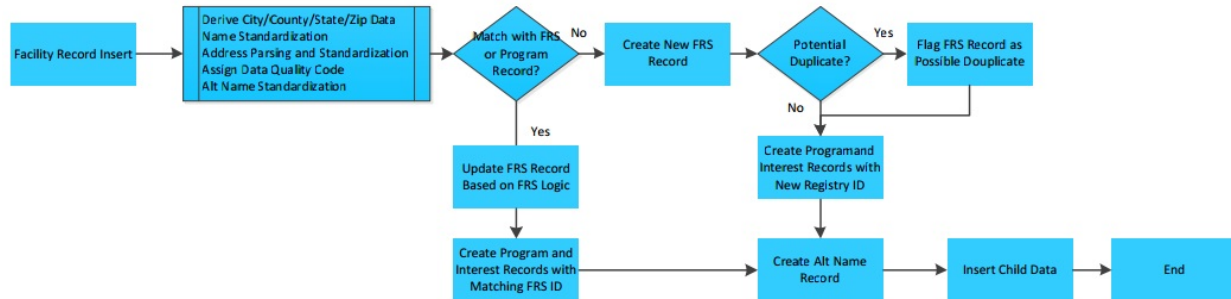
In the matching process, the incoming records are first standardized, parsed, and evaluated for data quality. These processes makes the data uniform and facilitates in the matching process, as it makes comparing the records easier and faster. FRS performs dynamic searches of the existing database using name and address matching algorithms, which results in a score, as shown below. If the record is matched with a FRS or program record with a score of 85 or above, the incoming record is linked to the FRS record and then goes through the update process. If the matching score for the incoming record is 84 or below, a new FRS record is created. If the score is between 35 and 84, FRS flags the record as a potential duplicate for a steward to review. If the score is 34 or below, no flags are raised and no linking takes places, it is treated as a completely new record in FRS.

The data elements are given scores and the sum of their scores determines whether the facility will be merged or flagged. The scoring is based on the following standardized data elements:

Scoring Table	
Street name and House Number	50
Facility name	25
City	15
County	5
Zip	5

In the three scenarios below, a match on the data elements shown would add up to a score of 85 and above, in which case the source record is linked with the FRS record.

Merge combinations		
Scenario 1	**Scenario 2**	**Scenario 3**
City	City	County
County	State Code	Zip
Zip Code	House Number	State
State Code	Street Name	House Number
House Number	Facility Name	Street Name
Street Name		Facility Name
Facility Name		

Manual Merging

Manual merging of duplicate facilities is done through the Facility Linkage Application (FLA). When the automated process detects close potential matches to facilities, but not exact duplicates, the facilities are flagged as possible duplicates. The possible duplicates then must be resolved by a data steward to confirm if these facilities are the same through research. If duplicates are present the data steward then merges the duplicate facilities, consolidating the multiple records into one FRS record. If through research it is deemed that the given facility is a new facility, and not a duplicate; then the data steward simply clears the flag. Facilities can also be merged directly by the data steward through FLA, if they are known to be duplicates without being flagged

Data Stewards

A primary data steward acts as a point of contact for regional data needs and data quality issues. Assists with the dissemination of information about FRS, FRS data improvement projects, and coordinates with State Data Stewards. The primary data stewards make data corrections in FLA, which are incorporated directly into the FRS production layer, and approve data corrections submitted by secondary or emergency response stewards within their region. Each EPA region is represented by a primary data steward; a list is available from the FRS website or by clicking here. Secondary data stewards consist of staff from regions, program offices, states and trusted partners not identified as primary stewards. Secondary data stewards are delegated responsibility by the primary stewards. These stewards recommend changes that are later approved by their regional primary data steward.

Merging through the FLA "Possible duplicates"

There are two ways to merge facilities in FLA

- Via the possible duplicates or resolve duplicates report
- Merging facilities not in the duplicates reports (if known, through other reports, searching FLA, etc.)

There are two ways to access facilities flagged as possible duplicates

1.) Through home page quick access under the "DQ Initiatives Overall" at the bottom of the FLA page. The "Review" button in the second box will return a report of all possible duplicates.
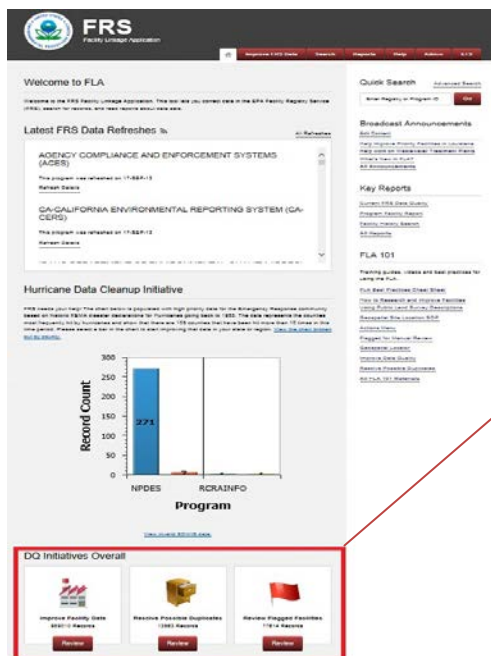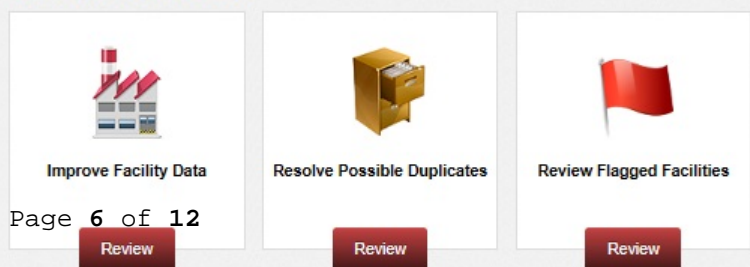
2.) Via the top navigation menu in the application, under the "Improve FRS Data" tab as the second link named "Possible Duplicates"
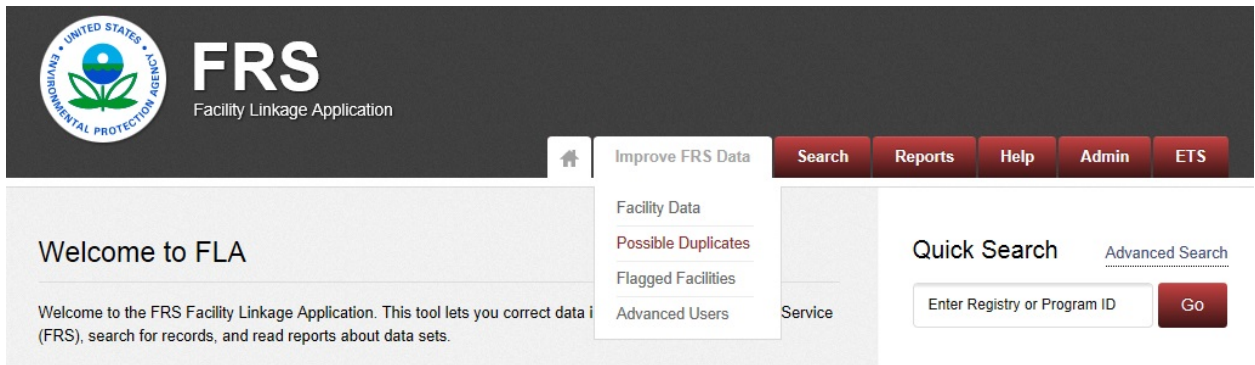
Both options will return a report of facilities that have possible duplicates in the entire country. The data steward can then choose to filter the report to narrow down the list to their desired preference.

Filtering reports:

- Click on the Actions drop down list and choose the "Filter" option
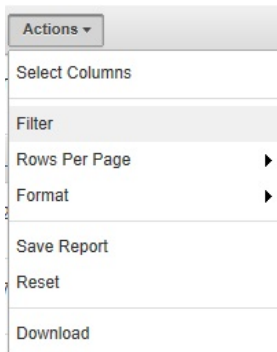
- A filter box will appear, allowing the data steward to choose multiple filters based on column, row, operator and expression.

For example if the data steward wanted to filter by the column "state". The data steward would choose state under the dropdown menu under column, equals (=) as the operator, and either type desired state or choose from the drop down list.

Once the dataset has been narrowed down to the desired aspects, the user selects the facility by clicking on the "Review" button next to the desired facility.

☐	Review	110055662607	1-DAY PAINT & BODY #127	1100 COMMERCIAL AVE.	OXNARD	CA	93033	VENTURA	1100 COMMERCIAL AVE, OXNARD, CA 93030 US	V	09

Clicking on the review button returns a list of the facility and its chosen possible duplicate(s).

	Registry ID	Primary Name	Location Address	City	State	Zip	County	Standardized Address	DQ Code
☑	110055662607	1-DAY PAINT & BODY #127	1100 COMMERCIAL AVE.	OXNARD	CA	93033	VENTURA	1100 COMMERCIAL AVE, OXNARD, CA 93030 US	V
☐	110013917707	1-DAY PAINT & BODY CENTERS INCORPORATED	1100 COMMERCIAL AVENUE	OXNARD	CA	93030	VENTURA	1100 COMMERCIAL AVE, OXNARD, CA 93030 US	V

1 - 2

Merge	Clear Flag	Reset	Done

Through research, the data steward chooses to either merge facilities or clear the flag. One or more facilities can be merged to selected facility if they are verified as the same facility.

Proper research includes looking at the facility detailed view(s) to look at associated data such as, alternate name(s), interests, SIC, NAICS codes and geospatial data to see if they are related. Search engines such as Google,

Bing along with map searches should also be employed in order to ensure that the facilities are the same. In our example, we can navigate to the 1-Day Paint and Body Centers webpage and search for locations. This confirms that the above facilities are duplicates and should be merged.

In order to merge the facilities click the box (to add a check) to the facility being merged and then click on the red "Merge" button. The data steward will be taken to the page shown below, which lists the facilities and its associated interests.

Merge	Registry ID	Primary Name	Location Address	City	State	Zip	County	Standardized Address	Update Date	Last Reported Date
⦿	110013917707	1-DAY PAINT & BODY CENTERS INCORPORATED	1100 COMMERCIAL AVENUE	OXNARD	CA	93030	VENTURA	1100 COMMERCIAL AVE, OXNARD, CA 93030 US	30-NOV-12	
	(NEI - NEICA1115625)	1-DAY PAINT & BODY CENTERS INC	1100 COMMERCIAL AVENUE	TORRANCE	CA	90501		TORRANCE, CA 90501 US	30-JAN-09	
	(EIS - 1372011)	1-DAY PAINT & BODY CENTERS INC	1100 COMMERCIAL AVENUE	TORRANCE	CA	90501	VENTURA	TORRANCE, CA 90501 US	30-NOV-12	
○	110055662607	1-DAY PAINT & BODY #127	1100 COMMERCIAL AVE.	OXNARD	CA	93033	VENTURA	1100 COMMERCIAL AVE, OXNARD, CA 93030 US	15-SEP-13	16-AUG-13
	(CA-CERS - 10200784)	1-DAY PAINT & BODY #127	1100 COMMERCIAL AVE.	OXNARD	CA	93033	VENTURA	1100 COMMERCIAL AVE, OXNARD, CA 93030 US	15-SEP-13	16-AUG-13

1 - 5

Comments

Choose Comments: [Duplicate Facilities ▾]

Additional Comments: [   ]

[**Merge**] [Merge and Update] [Reset] [Clear Flag]

Here the data steward will be choosing one record and FRS Registry ID to keep, and the environmental interests from the other facility records will be merged into the one that is retained.

After researching, if it is deemed that the facility data needs to be updated, the data steward can choose the "Merge and Update" button. This will allow the data steward to update the chosen facility before merging them.

Once the facility has been cleared or merged, a message will be displayed showing the data steward that it was successfully merged or if there was an error.

Merging known duplicates:

Another method for merging is manually selecting and merging facilities that are known to be duplicates. An example would be to merge two facilities that the data steward knows for a fact are duplicate, but have not made it to the possible duplicates report or have had the flag incorrectly removed. There are various reasons why a facility would not show up in the possible duplicates report, one example is if the naming is incorrect or address has been wrongly entered – In which case you would update the facility data first. The steps below explain how to merge facilities not in the "possible duplicates" report. Duplicates in the system can also be found and merged with facilities found to be the same while searching FLA.

For demonstration purposes we will use the following business:

1-Day Paint & Body # 102
Located at: 21801 S Western Ave, Torrance, CA, 90501
With the registry ID: 110055693076

The data steward already knows the facility has a duplicate. The duplicate facility name and information is:

1-Day Paint & Body Centers Inc
Located at: 21801 S Western Ave, Torrance, CA, 90501
With the registry ID: 110002694273

In order to merge these two facilities the data steward would first search for one of the facilities. This can be done two ways, by using the "Quick search" in the home page on the right side or through the more comprehensive "Search" tab.
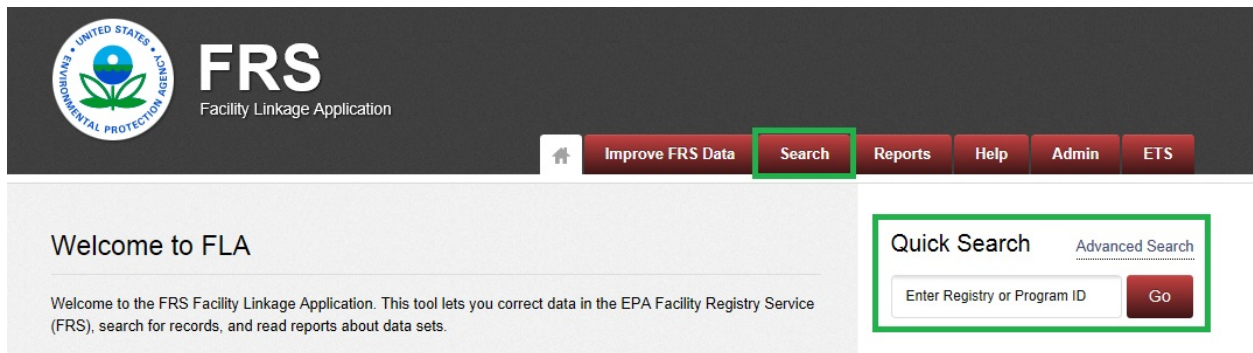
Once you search for the facility it will return the search result. The data steward would then click on the "Facility details" button next to the facility
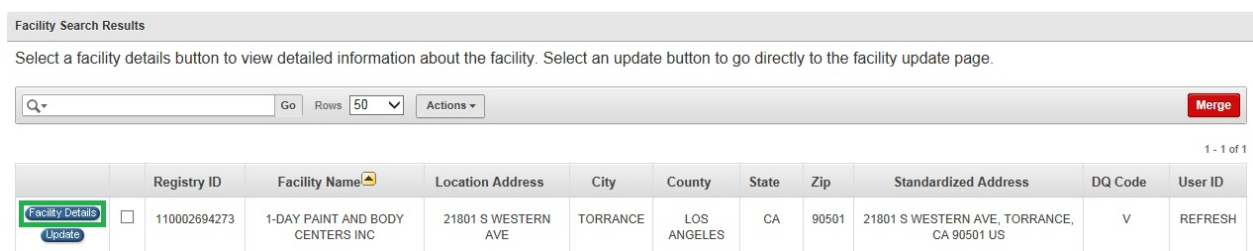
The facility detail page will show up, and then the data steward would click on the "Search for duplicates" button
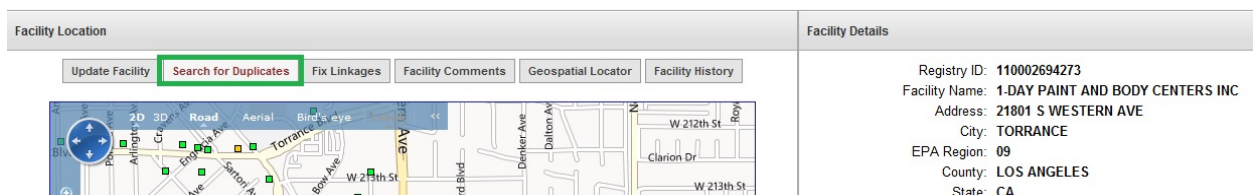
Then click on the Merge button (if there are options present, but the options do not include the facility the data steward is intending to merge. Click the merge button without selecting or checking any or the boxes for the given facilities)

Facility Matching Results

Analyze the duplicate candidates and place a check in the box beside any duplicate records.

- Select any registry IDs to merge by clicking the check box.
- Additional registry IDs can be added manually on the next screen after pushing the Merge button.
- Select the Done button to not merge any records and return to the previous page.

	Registry ID	Primary Name	Location Address	City	State	Zip	County	Standardized Address	DQ Code
☑	110002694273	1-DAY PAINT AND BODY CENTERS INC	21801 S WESTERN AVE	TORRANCE	CA	90501	LOS ANGELES	21801 S WESTERN AVE, TORRANCE, CA 90501 US	V

Merge **Reset** **Done**

If there are no duplicates found, or if you click on the merge button without selecting any proposed facilities. You will be brought to the next page. Where the data steward will click on the "Manually Add Facilities"

No duplicates found

- No duplicates have been found for this facility.
- Registry IDs to merge can be added manually on the next page by selecting the Manually Add Duplicates button.
- Otherwise, select the Return button.

Return **Manually Add Duplicates**

The next page will provide the data steward with the ability to add a registry ID you would like to merge with the facility. Click on the Add button.

Merge Records

Identify the record that you would like to keep as the FRS record.

- Select the record which retains the most history by first selecting the record that has the most linkages, second, by the oldest update date.
- Select Duplicate Facilities from the Choose Comments dropdown
- Enter any Additional Comments.
- Select the Merge button.

Registry ID: 110055693076 ✕ **Add**

The facility will be added and then the data steward would click on the merge button after selecting the radio button next to facility deemed to have more interests or historical data. Any desired comments can be added.

Merge	Registry ID	Primary Name	Location Address	City	State	Zip	County	Standardized Address	Update Date	Last Reported Date
⦿	110002694273	1-DAY PAINT AND BODY CENTERS INC	21801 S WESTERN AVE	TORRANCE	CA	90501	LOS ANGELES	21801 S WESTERN AVE, TORRANCE, CA 90501 US	08-AUG-10	
	(HWTS-DATAMART - CAD981400104)	1-DAY PAINT & BODY CENTERS, INC	21801 S WESTERN AVE	TORRANCE	CA	90501	LOS ANGELES	21801 S WESTERN AVE, TORRANCE, CA 90501 US	18-NOV-04	18-NOV-04
	(RCRAINFO - CAD981400104)	1-DAY PAINT AND BODY CENTERS INC	21801 S WESTERN AVE	TORRANCE	CA	90501	LOS ANGELES	21801 S WESTERN AVE, TORRANCE, CA 90501 US	08-AUG-10	
○	110055693076	1-DAY PAINT & BODY #102	21801 S WESTERN AVE	TORRANCE	CA	90501	LOS ANGELES	21801 S WESTERN AVE, TORRANCE, CA 90501 US	15-SEP-13	31-JUL-13
	(CA-CERS - 10131295)	1-DAY PAINT & BODY #102	21801 S WESTERN AVE	TORRANCE	CA	90501	LOS ANGELES	21801 S WESTERN AVE, TORRANCE, CA 90501 US	15-SEP-13	31-JUL-13

1 - 5

Comments

Choose Comments: [Duplicate Facilities ▾]

Additional Comments: [   ]

[Merge] [Merge and Update] [Reset]

The facility will be successfully merged and a message should display on the facility details page, where the data steward will be redirected.

✓ Record(s) have been successfully merged.   ☒

Facility Location

[Update Facility] [Search for Duplicates] [Fix Linkages] [Facility Comments] [Geospatial Locator] [Facility History]

Facility Details

Registry ID:	110002694273
Facility Name:	1-DAY PAINT AND BODY CENTERS INC
Address:	21801 S WESTERN AVE
City:	TORRANCE
EPA Region:	09
County:	LOS ANGELES
State:	CA
Zip:	90501
Legislative District Number:	4
Federal Facility Code:	N
Federal Agency Code:	
Federal Agency Name:	
Tribal Land Code:	N
Reviewed By:	AMAR_SAHASRABUDHE
Last Updated:	18-SEP-13
Review Reason:	
Data Quality Code:	V
Location Description:	
Congressional District:	
U.S. Mexico Border:	
HUC:	